

Copyright
by
Jennifer Marie Raye
2017

The Report Committee for Jennifer Marie Raye
Certifies that this is the approved version of the following report:

HathiTrust Digital Library and Fair Use

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Philip Doty

Stephanie Towery Swenson

HathiTrust Digital Library and Fair Use

by

Jennifer Marie Raye, B.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Information Studies

The University of Texas at Austin

May 2017

Acknowledgements

This report would not have come to fruition without the incredible love and support of my family, friends, and advisors. To my parents, you have always supported my education, even when I wanted to change high schools, go away to college, and then move to Texas. I know leaving has not been easy for you, but thank for you loving, supporting, and continuing to send me care packages. I'm so proud to be your daughter. Words cannot describe how grateful I am to have you as parents.

To Christian, you took a leap of faith and moved with me from College Station to Austin. You loved and supported me when I needed a ride home from the library at 2 a.m., for being a voice of reason when my crazy ideas were just a little *too* crazy, and letting me be my true self. Thank you for agreeing to go on this adventure to Austin and I can't wait to see what the next chapter has in store for us.

To my advisors Dr. Philip Doty and Stephanie Swenson Towery, thank you for agreeing to be part of my committee. Dr. Doty, you pushed me to be a better, stronger writer. Thank you for providing your expertise on this report among the million things you already have to do. Stephanie, your class gave me the opportunity to explore the intersection of law and libraries. I'm so glad I took your class and the path I'm on because of your class.

Thank you to the members of HathiTrust who participated in my interviews. Your insight has brought more clarity and support to this report. Thank you for taking the time to be a part of this report.

To Sophie, my hedgehog, you stayed up with me till all hours of the night during my undergraduate and graduate school career. You made the journey from Michigan to Texas like a champ. Thank you for keeping me company during my all-nighters and being my little water balloon with legs. Patch, my one-eyed papas, you and Soph are the best fur kids ever. Thank you for being the laziest, best dog anyone could ask for.

Thank you to everyone who has been a part of my life, your presence, has shaped me into the woman I am today. For this, I am forever grateful.

Abstract

HathiTrust Digital Library and Fair Use

Jennifer Marie Raye, MSInfoStds

The University of Texas at Austin, 2017

Supervisor: Philip Doty

Preservation and dissemination of information is the cornerstone of most libraries. With the onset of the digital age, libraries of all sizes began scanning and cataloging their older documents and pictures, making them available on the Internet. In 2004, a consortium of 13 university libraries partnered with Google digitized millions of unique titles, growing into the HathiTrust Digital Library. HathiTrust developed with the goal of addressing the brittle books problem plagued by libraries and archives, and evolved into providing access to these digitized materials. By 2011, the Authors Guild filed lawsuits on behalf of its members against HathiTrust and Google, separately for copyright infringement and unauthorized reproduction. Both HathiTrust and Google maintained a fair use defense.

HathiTrust effectively demonstrated their database and its availability to print-disabled patrons was highly transformative. By digitizing print, analog works, Google was able to create a searchable database of materials. In addition to the search function, libraries were

able to use their digitized copy for their print-disabled patrons. The District Court and Court of Appeals upheld HathiTrust and Google's fair use defense, citing the highly transformative use contributing the most to the fair use balance.

The HathiTrust and Google decisions demonstrate how fair use applies in the context of digitization projects within libraries. The HathiTrust decision specifically, serves as a guide for libraries looking to digitize their analog works collections. By following characteristics described in the court opinions, libraries can maintain copyright law compliance. For example, by restricting reproductions of copyrighted materials for preservation and transformative uses, libraries can exercise fair use practices. Through an interview with a member of HathiTrust, more considerations are provided for libraries, they can decide to follow a toolkit provided by HathiTrust, join HathiTrust, or find a combination of options that best fits their institution. As seen by the District Court judge as one of the most important reasons for upholding the HathiTrust, access for the print-disabled communities provides more than just access for a marginalized community, it provides a public good for society. By following some of these conditions, libraries can feel confident going through with a digitization project.

Table of Contents

List of Tables	xi
List of Figures	xii
INTRODUCTION.....	1
HATHI [®] TRUST DIGITAL LIBRARY.....	6
Searching and Accessibility	7
Searching in HathiTrust Digital Library	7
Accessibility for Print-Disabled Patrons	11
Technology, Standards, and Specifications	12
Access Determination for Digital Objects	12
Shibboleth Login.....	13
Rights Database	13
HathiTrust’s Metadata Structure.....	14
Conclusion.....	17
HOW HATHI [®] TRUST AND GOOGLE WON	19
A Brief History of Fair Use	24
Background to <i>Authors Guild v. HathiTrust</i> (2012) and <i>Authors Guild v. Google</i> (2013).....	26
The Fair Use Defense in the District Court and Court of Appeals Cases	26
The Purpose and Character of the Use	29
The <i>Authors Guild v. HathiTrust</i> Cases	29
Authors Guild v. HathiTrust, 2012.....	29
Authors Guild v. HathiTrust, 2014.....	31
The <i>Authors Guild v. Google Cases</i>	32
Authors Guild v. Google, 2013	32
Authors Guild v. Google, 2015	32

Nature of the Copyrighted Works.....	33
The <i>Authors Guild v. HathiTrust</i> Cases	33
Authors Guild v. HathiTrust, 2012.....	33
Authors Guild v. HathiTrust, 2014.....	34
The <i>Authors Guild v. Google</i> Cases	34
Authors Guild v. Google, 2013	34
Authors Guild v. Google, 2015	35
The Amount and Substantiality of the Portion Used.....	35
The <i>Authors Guild v. HathiTrust</i> Cases	35
Authors Guild v. HathiTrust, 2012.....	35
Authors Guild v. HathiTrust, 2014.....	36
The <i>Authors Guild v. Google</i> Cases	36
Authors Guild v. Google, 2013	36
Authors Guild v. Google, 2015	37
Google’s Snippet View	37
The Effect on the Potential Market or Value of the Copyrighted Work	38
The <i>Authors Guild v. HathiTrust</i> Cases	39
Authors Guild v. HathiTrust, 2012.....	39
Authors Guild v. HathiTrust, 2014.....	40
The <i>Authors Guild v. Google</i> Cases	40
Authors Guild v. Google, 2013	40
Authors Guild v. Google, 2015	41
The Overall Assessment	41
The <i>Authors Guild v. HathiTrust</i> Cases	42
The <i>Authors Guild v. Google</i> Cases	43

Conclusion.....	44
CONSIDERATIONS FOR MASS DIGITIZATION PROJECTS WITHIN LIBRARIES	45
Essential Digital Library Characteristics for a Fair Use Defense.....	45
The Transformative Nature of the Digital Library	46
Changes to HathiTrust after the Litigation	47
Joining HathiTrust and the Copyright Review Management System Toolkit	47
Reproductions for Use and Preservation.....	48
Access for Print-Disabled Patrons.....	50
Conclusion.....	52
CONCLUSION	53
REFERENCES	56

List of Tables

Table 1. Plaintiffs and defendants in the four cases.....	22
Table 2. Key points of the District Court cases.....	27
Table 3. Key points of the Court of Appeals cases.....	28

List of Figures

Figure 1. Search results for “Google” in the HathiTrust Digital Library.	9
Figure 2. “Google hacks” catalog record.	10
Figure 3. “About this Book” includes the rights information related to this search result.....	10

INTRODUCTION

Preservation and dissemination of information are cornerstones of library function. The digital age created new opportunities for libraries, archives, and museums to expand beyond print materials collection and circulation. Large-scale digitization efforts have been on-going since the 1970's, but were often faced with financial, technological, and social limitations. Now, with current digitization technologies, institutions of all sizes are scanning and digitizing their analog books, documents, pictures, and various other materials. Large-scale or mass digitization projects ("MDP") conducted by libraries, museums, and archives expand access to the library's materials by moving printed items from the library's shelf to users across the world. Whether a library is undertaking a digitization project to preserve materials for future generations or to make them available to a wider audience, these digitization initiatives are creating functional and findable digital objects from previously isolated works.

In December 2004, a partnership developed to expand these digitization initiatives on a nation-wide scale. By 2008, a 12- university library consortium partnered with Google, Inc. began digitizing millions of unique titles, "launching a new era of large-scale digitization previously not feasible or affordable" (Bellardo, 2006, p. 20). This large-scale digitization initiative evolved into HathiTrust (www.hathitrust.org). The California Digital Library, Carnegie Mellon's Million Books Project, and others developed their own MDPs. HathiTrust, however, was partnered with Google, a for-profit, commercial corporation. When copyrighted works are made available on the Internet or within a publicly accessible database, copyright holders are bound to take notice, especially when Google was funding and conducting the digitization process. The unique titles digitized and accessible through

HathiTrust include works currently under U.S. copyright protection and within the public domain. HathiTrust has significant copyright implications, not just for Google and HDL, but for libraries across the U.S. conducting digitization projects. Critics of the HDL include authors, author advocacy groups, and publishing companies, claiming copyright infringement. So much so, in 2011, the Authors Guild, a professional organization that advocates for the rights of writers related to contracts and copyrights, and several individual authors filed lawsuits against HathiTrust and Google for copyright infringement and unauthorized reproduction. In both lawsuits, plaintiffs alleged that Google's copying of the entire texts of copyrighted works was infringement, regardless of its use as a searchable database. HathiTrust and Google cited fair use, effectively demonstrating that their database and its availability to print-disabled patrons is highly transformative.

This report aims to understand how a mass digitization project needs to be developed in order to maintain compliance with current copyright law, particularly the fair use clause in copyright law (17 U.S.C. § 107). By examining the infrastructure of HathiTrust as well as the District Court and Court of Appeals rulings for both HathiTrust and Google cases, researchers and library professionals will have a better understanding of how to develop MDPs proactively with copyright to advance the public's access to information. The HathiTrust and Google decisions demonstrate how fair use applies in the context of digitization projects by libraries. Additionally, by providing first hand knowledge through an interview of a highly placed member of the HathiTrust staff, we gain more insight into the mission and implications of HathiTrust. The first section of this report surveys the current infrastructure and policies of HathiTrust Digital Library ("HDL") available on their website. HDL's infrastructure is particularly important because the digital library uses a specific

framework when making digital objects available. The framework includes functions that ultimately determine if a user's search results are freely accessible or the results are restricted, copyrighted works. These functions include the search engine tool, the PageTurner application, access determination for objects, and the rights database. Metadata can provide information regarding the digital object's current copyright status and the potential copyright term expiration date. Each digital object in HDL contains a metadata record, information including, but not limited to the digital object's author/creator, date of creation, publication date, language, physical description, and where to locate a print version. These tools work together to determine if a user can download a copy of their searched work.

The next section of this report focuses on copyright implications of making a mass digitization project available online. The openness and sharing ethos of the Internet disrupted an established business model for copyrighted works. The Internet created new case precedent in the intellectual property field through cases like *Sega Enters. Ltd. v. MAPHIA*¹, *A&M Records, Inc. v. Napster, Inc.*², and *Kelly v. Arriba Soft Corp.*³. The Google and Hathitrust digital library projects raised significant copyright concerns in comparison to its predecessors, such as the previously mentioned California Digital Library and Million Books Project, which primarily dealt in public domain materials. This section explains the various elements of copyright law implicated by MDPs of copyrighted and public domain works, focusing heavily on the statutory exemption of fair use. More specifically, this section will

¹ *Sega Enters. Ltd. v. MAPHIA*, 948 F. Supp. 923 (N.D. Cal. 1996), fair use when MAPHIA, an online bulletin board, enabled users to upload and download Sega Enterprises, Ltd.'s copyright-protected video games for home use.

² *A&M Records, Inc. v. Napster, Inc.*, 239 F. 3d 1004 (9th Cir. 2001), not fair use when Napster, Inc. designed and operated a peer-to-peer file-sharing network, enabling users to search, upload, and download audio recordings stored in digital file format on their own or other's computers.

³ *Kelly v. Arriba Soft Corp.*, 336 F. 3d 811 (9th Cir. 2003), fair use when Arriba Soft Corp.'s search engine crawled the Internet search for images that it copied and then generated as smaller, lower-resolution thumbnail copies for display on a search results page.

discuss the current structure of the judicial system, requirements for proving a copyright infringement lawsuit, and the background of the Southern District Court of New York's *Authors Guild v. HathiTrust* (2012)⁴ and *Authors Guild v. Google, Inc.* (2013)⁵. Next, I will discuss fair use's historical development from case law to a codified statute in the Copyright Act of 1976. Additionally, a modern evaluation of a fair use defense provides foundational support for the case discussion. Only then can we evaluate fair use's application in HathiTrust and Google's District Court and Court of Appeals cases. The opinions by the District Court and the Court of Appeals for the HathiTrust and Google Books cases are of the utmost importance, because they are the first instance of litigation against a mass digitization project. Litigation involving the Internet, digitization, and fair use had occurred earlier, of course. These cases, however, were the first instances of using copyrighted materials in a transformative fair use—a searchable database. The opinions provide insight into the legal system's understanding of fair use and HDL, and into how mass digitalization projects can function within the framework of fair use.

Finally, after discussing the components of HDL and the specific insights gained from the fair use assessments conducted by the District Court and the Court of Appeals for *Authors Guild v. HathiTrust* and *Authors Guild v. Google*, I provide suggestions for libraries considering digitizing their collections. Suggestions provide essential characteristics the digital library must have in order to comply with the fair use characteristics upheld through the court system. By following the Copyright Review Management System toolkit, as mentioned by a HathiTrust staff member, libraries can identify and provide access to their public domain collections. Next, I will discuss how the transformative aspects contribute to fair use, such as

⁴ *Authors Guild v. HathiTrust*, 902 F. Supp. 2d 445 (S.D.N.Y. 2012)

⁵ *Authors Guild v. Google, Inc.*, 954 F. Supp. 2d 282 (S.D.N.Y. 2013)

how the works are presented to the user. If a library would like more specific guidelines and protections, the library does not need to rely on fair use, but can investigate Section 108 of the copyright law. Depending on the library's procedures, the library may qualify for more protections. Additionally, the University of Michigan has provided a Copyright Review Management System toolkit. This publication describes the efforts to conduct copyright review of books involved in digitization projects. The toolkit provides additional information to help others "responsibly identify and provide meaningful access to the public domain collections"(Levine, Adler, Bonfiglio, Eden, & Hall, 2016). Overall, HDL provides a public good. HathiTrust's transformative use goes beyond just a library search engine, but provides access to materials for underserved populations, such as print-disabled communities, and preserves their digital objects for future generations.

HATHITRUST DIGITAL LIBRARY

HathiTrust is a partnership of major research institutions and libraries working to ensure that the cultural record is preserved and accessible long into the future. HathiTrust's mission is similar to that of most libraries, to "contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge" (HathiTrust, 2017b). HathiTrust's main goal is to build a reliable and increasingly comprehensive co-owned and co-managed digital archive of library materials converted from the print collections of its member institutions. "In the beginning, the brittle book problem was the biggest issue our library faced," said a highly placed HathiTrust staff member in a personal interview, "we never would have had the resources to make our collections accessible on this scale." In addition to developing a digital archive, providing access to public domain and in-copyright materials to print-disabled patrons has been a mission of any library, and HathiTrust is no exception.

Other goals of the HathiTrust Digital Library include developing a cost-effective infrastructure for digital content, developing partnerships to ensure preservation of digital objects, reducing long-term costs of storage and care of print collections, and sustaining HathiTrust as a public good while maintaining services that benefit member institutions. HathiTrust's goals coincide with many library institutions' mission and goals to promote, disseminate, and preserve information. This section of the report describes the current HDL infrastructure, focusing on the search capabilities and accessibility of the digital library. Investigating further the infrastructure further, we will see how the technology and standards developed by HathiTrust, including the Rights Database, the login requirements, and the metadata for each digital object contribute to a user's (in) ability to access a particular work.

Since its development in 2008, HathiTrust has expanded to include over 120 library partners and is open to institutional membership worldwide. HathiTrust is comprised of the digital library, their collaborative programs, and their research center. The digital library is the largest component of HathiTrust. It is a digital preservation repository which maintains the digital collection of partnered institutions. As of February 2017, HathiTrust has digitized almost 667 terabytes of files, including 15 million volumes, which includes text, photos, and other materials, over 7 million book titles, and over 5 billion pages (HathiTrust, 2017c).

Searching and Accessibility


In order to evolve HathiTrust from a strictly preservation archive to a publicly accessible database, HDL developed strong search capabilities. Without a well-defined search function available, patrons are not able to use a digital library's materials. HathiTrust's search functions are largely based on the bibliographic information about its digital objects. The bibliographic information includes title, author, subject, ISBN, publisher, and year of publication. HathiTrust allows searching through its homepage or through its partnership with Online Computer Library Center (OCLC)'s catalog.

SEARCHING IN HATHI TRUST DIGITAL LIBRARY

In addition to the homepage's search function, HathiTrust offers full-text search across the digital repository, which includes both public domain and copyrighted works. Partnered institutions are able to load HathiTrust's bibliographic records into their home library catalogs and search engines. A bibliographic search can produce many results, but a filtered, advanced search for both public domain and copyrighted works can be achieved with a robust global search engine. After a search is complete, the user sees results of the search. It

is not indicated how the results are ranked. The results, however, are divided into two tabs, “All Items” and “Full View” (HathiTrust, 2016). The “All Items” tab includes “Limited (Search-only)” and full view records. The “Full View” tab includes records available for full view and downloads. For example in Figure 1, when “Google” is searched, we see 14,595,292 items in *Full-Text + All Fields*. “Google hacks: 100 industrial strength tips and tricks,” the first search result allows access to only the catalog record (Figure 2) and states, “This item is not available online (Limited-search only) due to copyright restrictions” (HathiTrust, 2016). The catalog record displays rights information for the selected item, as seen in Figure 3. Clicking “Limited (Search-only)” allows the user to search inside the text to find the frequency and page numbers of specific words and phrases. Next to this search box, the record is linked with OCLC’s WorldCat database, allowing the user to find a local “physical” copy of the work.

[Home](#)
[About](#)
[Collections](#)
[Help](#)
[Feedback](#)



FULL-TEXT

CATALOG

Advanced full-text search

Search tips

Full view only

LOG IN

Refine Results

Subject

- [United States](#) (355,560)
- [Agriculture](#) (174,486)
- [China](#) (147,011)
- [Science](#) (139,643)
- [Dissertations, Academic](#) (130,284)
- [more...](#)

Author

- [United States. Bureau of the Census.](#) (31,946)
- [Geological Survey \(U.S.\)](#) (25,630)
- [U.S. Atomic Energy Commission.](#) (18,507)
- [Shakespeare, William, 1564-1616.](#) (15,366)
- [United States. Bureau of Labor Statistics.](#) (14,929)
- [more...](#)

Language

- [English](#) (7,558,189)
- [German](#) (1,323,319)
- [French](#) (1,053,809)
- [Spanish](#) (827,394)
- [Chinese](#) (522,775)
- [more...](#)

Place of Publication

- [United States](#) (4,817,091)

Search Results: 14,592,292 items found for **Google** in *Full-Text + All Fields*

All Items (14,592,292)

Full View (5,356,026)


25 per page

1 2 3 4 5 6 7 8 ... 800 Next

Select all on page

Select Collection


Add Selected

☐


Google hacks : 100 industrial strength tips and tricks / Tara Calishain and Rael Dornfest.

by Calishain, Tara.
Published 2003


[Catalog Record](#) [Limited \(search-only\)](#)

☐


Google pocket guide / Tara Calishain, Rael Dornfest and DJ Adams

by Calishain, Tara
Published 2003


[Catalog Record](#) [Limited \(search-only\)](#)

☐


Die Google Falle : die unkontrollierte Weltmacht im Internet / Gerald Reischl.

by Reischl, Gerald.
Published 2008

[Catalog Record](#) [Limited \(search-only\)](#)

☐


Hacking Google Maps and Google Earth / Martin C. Brown.

by Brown, Martin C.
Published 2006

Figure 1. Search results for “Google” in the HathiTrust Digital Library.

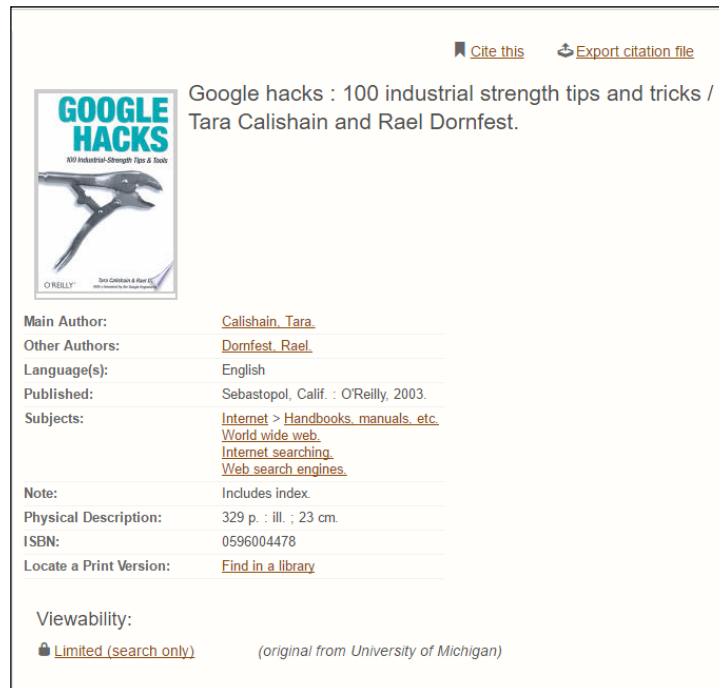


Figure 2. "Google hacks" catalog record.

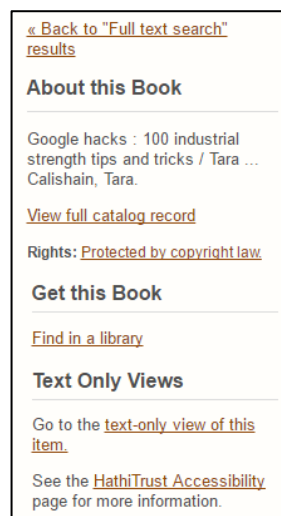


Figure 3. "About this Book" includes the rights information related to this search result.

The results include public domain and unrestricted items and in-copyright (restricted) items. Users can search within individual volumes, when they are confirmed to be in the public domain, and read through the page-turner application. The page-turner application renders the individual files as native browser files, PDFs, and optical character recognition (OCR) text. The HathiTrust PageTurner is a publicly accessible application. When a search is

entered, the PageTurner application determines the Final Access Status of “allow” or “deny.” The application determines the status based on several data sources, including login authentication systems, originating IP address, HathiTrust Rights Database, HathiTrust Print Holdings Database, GeoIP Database, and HathiTrust Institution Orphan Access Agreement Database. The user’s IP address is particularly important for determining “allow” or “deny” access. IP, or Internet protocol, is a unique string of numbers that identifies each computer using the Internet protocol to communicate over a network. A specific computer’s IP address identifies the geographical location of an Internet user. If a HathiTrust user is accessing the library from Austin, Texas, HDL cannot “allow” access to in-copyright materials, except those in the library being used, as outlined by U.S. Copyright law. All data sources must agree the final status is “allow” in order for the full-view to be available for a specific work.

ACCESSIBILITY FOR PRINT-DISABLED PATRONS

As discussed earlier, one of HathiTrust’s main goals is to disseminate information to print-disabled users. The University of Michigan was an early leader in digitizing works for print-disabled patrons. Through a designated proxy, eligible, certified print-disabled users affiliated with the University of Michigan have lawful, full-view access to public domain and in-copyright works. HathiTrust also developed a version of the print-disabled access that other libraries can use. Currently, the University of Michigan is the only HathiTrust partner with print-disabled patron access available; however, other institutions are developing their own certified print-disabled user policies to provide access to qualified users. Eligible patrons at HathiTrust institutions can receive special lawful access to in-copyright materials, through the PageTurner application, which contains special markup information for print-disabled patrons. The application currently meets Section 508 requirements (29 U.S.C. § 794 (d))

outlining compliance guidelines for making electronic and information technology accessible to people with disabilities.

Technology, Standards, and Specifications

HathiTrust's technological infrastructure is intended to facilitate search capabilities and act as a preservation and storage center of the digital objects. In order for HathiTrust to continue providing persistent and high-availability storage for deposited files, HathiTrust has developed a technological infrastructure designed to prevent loss of data. HathiTrust has two synchronized storage facilities, located in Ann Arbor, MI, and Indianapolis, IN. The infrastructure includes an encrypted tape backup with six months of previous-version retention. All storage centers are physically secure, with accessibility to only authorized IT personnel. These high standards contributed to the eventual fair use decisions discussed in the latter sections of this report.

ACCESS DETERMINATION FOR DIGITAL OBJECTS

HathiTrust must determine who has access to digital objects before the objects are available in the search function. Before the PageTurner application provides the user with the digital copy a work, the application must first read the status of an object. The "final access status" allows or denies access to the work. The allow status is based the several data sources mentioned earlier, including the bibliographic metadata of the digital object as recorded in the Rights Database. The Rights Database provides the source values and rights attributes for digital objects. In order to reach final access status, the status is provided from the "user types and rights" matrix. The user types include ordinary users, print-disabled users, users within a library building, University of Michigan affiliates, and HathiTrust affiliates. The

access status is reported by the Rights Database and the rights attributes, which include the categories (HathiTrust, 2017a):

- the public domain,
- in-copyright, undetermined copyright status,
- available world-wide,
- unavailable to anyone,
- public domain determined only in the U.S. and determined outside of the U.S.,
- works that carry a creative commons license,
- works that maybe protected by U.S. copyright law and require permission

(HathiTrust, 2017a).

When final access status allows, the user can download the available PDF.

SHIBBOLETH LOGIN

HathiTrust implements Shibboleth, an authentication mechanism for institutions that allows specialized services for patrons affiliated with partner institutions. The Shibboleth login grants users full-PDF download of public domain works, facilitated access to the Collection Builder, special access for users who have a print disability, and access to works held in print by partner institutions that are missing from the home institution. In order to receive these particular services, the HathiTrust partner institution (specifically the University of Michigan) provides login information for qualified, print-disabled patrons.

Rights Database

The Rights Database provides PageTurner application information about whether the user can download or print the PDF. The repository must store and track rights information about each digitized object in HathiTrust. In order to comply with copyright law in the U.S.,

libraries are making efforts to determine and express copyright status information for all of their holdings. When developing a digital repository for works in and out of copyright, some challenges include: organizing the rights information properly, ensuring accuracy in the semantics of rights, and tracking the changes to rights information over time. As mentioned earlier, one of the components of the digital object's access is the bibliographic information for the digital object, such as if the original object is a U.S. federal government work, the original object's county and date of publication. HathiTrust organizes rights based on the copyright status of the work. The database is divided into public domain and in-copyright works, which protects against errors in metadata changes, reduces errors in frequent updates to multiple digital objects, and ensures accuracy in legal terms. These attributes determine access to digital objects for users. The Rights Database is continually adding more attributes, such as if the copyright holder granted access later or if the copyright holder has added a Creative Commons license, for more complete digital object records.

HathiTrust's Metadata Structure

Metadata is particularly important for digital objects. HathiTrust stores and tracks rights information for each digitized item it holds. Metadata provides information regarding the digital object's record. Metadata can be used to identify items of similar type, such as genres of literary works, and can identify and track records specific to a contributor. HathiTrust requires as much information as possible when creating the metadata records for each object. HathiTrust's inclusion of Library of Congress Subject Headings creates a uniquely effective means of searching them higher recall and precision (Duffy, 2013, p. 5). HathiTrust utilizes MARC, a metadata standard used by libraries, for bibliographic information. Because MARC format cannot account for storage of rights information, HathiTrust uses a rights database.

Metadata is critical for determining copyright status of a digital object. Determining if an object is under copyright relies on the original object's publication/creation data, the location where the original object was published/created, and the author's personal information; these elements are considered the bibliographic information. The author's personal information is critical for copyright status determination because the current term of most works under a copyright protection is the life of the author plus 70 years. The bibliographic information predominantly determines the copyright status; copyright exceptions can occur, however. Many copyright-related access questions can be resolved by the bibliographic information alone, because U.S. federal government documents, country of publication, and publication date are some of the main criteria for determining the copyright term, if any. The country of origin and publication date are the first criteria used to determine the copyright status of a work. HathiTrust requires, when known, the publication date, the publication place, the material type description, the author, and further metadata information related to U.S. and non-U.S. publications.

For U.S. publications, the bibliographic metadata held by HathiTrust determines copyright status. If the U.S. federal government created the work, or the work was published in the U.S. prior to 1923, the work is in the public domain and can be used without restrictions to any user. The public domain is a collective term referring to creative materials that are not protected by copyright law. Public domain works are available to anyone to use without obtaining permission from the copyright holder (Stanford University Libraries, n.d.).

If a work does not meet either qualification, the work is protected by copyright, and HathiTrust restricts access as appropriate. Some works may qualify for unrestricted access

after a HathiTrust staff member reviews the bibliographic information. The copyright status of a work refers to the copyright law during the time of publication. Works published before 1964 in the U.S. are automatically in the public domain if a copyright renewal was never filed with the U.S. Copyright Office during the 28th year after publication. Also, works published prior to 1978 required copyright notice formalities. If the work did not include the word “Copyright” or a © and the name of the copyright owner, the work would enter the public domain. After researching records in the U.S. Copyright Office if a renewal was or was not filed or the original work did or did not contain a copyright notice, the metadata record should be updated to reflect this information (Stanford University Libraries, n.d.). If an author published a work after 1923, and filed appropriate formal notices and renewals, the work is feasibly under copyright protection. HathiTrust maintains the rights records and can add additional metadata attributes to their objects’ records when necessary.

For works created outside of the U.S., additional metadata attributes are needed. For non-U.S. works published from 1870 to 1923, the location of the user is necessary because the copyright status of a work can vary from county to county. Users with a U.S. IP address are granted access to non-U.S. works published from 1870 to 1923 because works published prior to 1923 are in the public domain for U.S. users. Due to variations in copyright law in other countries, however, some works may be still under copyright protection, restricting non-U.S. IP address users from accessing copyrighted material. Non-U.S. works published after 1923 may be under copyright, and further examination of a work’s bibliographic information is necessary to determine if the work is unrestricted for the user.

Conclusion

In this section, we identified and examined the components of HDL implicated in copyright law. The access infrastructure of HathiTrust has been designed around the concept of copyright terms and who can gain access to a work under the various exemptions allowed under copyright law. By examining the components of HDL's infrastructure we begin to see the specific framework HDL has developed and uses when making digital objects available. The search function provides users with results that are copyright protected, thus restricted to a search within function or they are full-view and available for download, at no time is an unauthorized user allowed full-view access to copyrighted works. The digital library uses various data source points, such as the Rights Database, origins of the user's IP address, the login authentication, and the metadata to determine if a work is available full view or limited access. Authorized, print-disabled patrons, through a proxy, are given access to in-copyright materials. This framework includes functions that ultimately determine if a user's search results are freely accessible or the results are restricted, copyrighted works. These characteristics of HathiTrust Digital Library will become important factors in the following discussion of fair use cases.

Regardless of the framework and infrastructure employed and policies adopted by HathiTrust and its partnered institutions, the next section will demonstrate how tensions began to develop between copyright holders and HathiTrust's Digital Library. In 2011, Authors Guild and several individual authors filed lawsuits against HathiTrust and Google for copyright infringement and unauthorized reproduction. Before describing and analyzing the four relevant cases (the District Court and Court of Appeals decisions for *Authors Guild v. HathiTrust* (2012) and *Authors Guild v. HathiTrust* (2014) and the District Court and Court

of Appeals decisions for *Authors Guild v. Google, Inc.* (2013) and *Authors Guild v. Google, Inc.* (2015)), I will provide information necessary for understanding a fair use defense in copyright litigation at the federal level. This is necessary to fully understand the weight of the opinions in the context of mass digitization projects conducted by libraries.

First, I will briefly outline copyright litigation procedure, such as filing in a specific federal jurisdiction, current copyright law and requirements for a plaintiff to show misappropriation and determining a defendant's available affirmative defenses. In order to evaluate the courts' findings of the fair use defense in the four cases, I will begin with a brief history of fair use and the statutory description of fair use. The largest portion of this section is dedicated to analyzing the four factors of fair use and how HathiTrust's application of fair use was evaluated in the case opinions. This section concludes with the overall assessment of fair use for each of the four court cases. By understanding the legal framework the case opinions have provided, we will be able to provide considerations libraries may wish to implement for their own digitization projects.

HOW HATHITRUST AND GOOGLE WON

By 2011, Authors Guild, a professional organization that primarily advocates on behalf of writers in the context of authors' rights and copyright claims, and several individual authors filed lawsuits against Google and HathiTrust for systematic digitization of copyrighted materials without authorization. Before describing and analyzing these lawsuits, however, I need to provide summary information for filing federal cases, litigation procedures for plaintiffs and defendants, and what rights are afforded to copyright holders.

In the United States, copyright law falls within federal jurisdiction. The U.S. has 94 active district courts, including one in the District of Columbia and Puerto Rico. District courts resolve legal disputes by applying legal precedent and principles to determine which party will prevail. Trial courts include a judge who tries the case and a jury that decides the case. After the proceedings have concluded, the losing party can appeal to the local U.S. Appellate Court. The 94 district courts are organized into 12 regional circuits. The appellate court's task is to determine whether the law was correctly applied in the district trial court. Appellate courts do not use a jury. After an appellate case has concluded, the losing party may file a writ of certiorari requesting the U.S. Supreme Court to hear the case. The Supreme Court, however, has almost complete discretion to choose the cases it will hear. According to a C-SPAN interview, the U.S. Supreme Court agrees to hear about one percent of all petitions it receives (Supreme Court of the United States, n.d.). The U.S. Supreme Court is the highest court in the U.S. and delivers the final judgment on a case.

According to the Annual Report of the Director, just over 5,000 copyright cases were filed in the 94 U.S. District Courts in 2015 (United States Courts, 2015). In the U.S., copyright cases

often occur in the Southern District Court of New York or the Northern District of California, because New York City and Los Angeles, respectively, fall within these districts. In geographic areas with concentrations of entertainment, audio-visual, and software industries, these district courts often see copyright cases and have developed their own tests and case law precedent. If the losing party of a Southern District of New York District Court appeals the court's decision, the Second Circuit Court of Appeals will hear the case. Both *Authors Guild v. HathiTrust* (2012) and *Authors Guild v. Google* (2013) were under the jurisdiction of the Southern District of New York due to the Authors Guild's headquarters in New York City.

When a plaintiff files a copyright lawsuit, the plaintiff must prove the infringing use violated their exclusive rights under §106 of the Copyright Act. In order to do so, the plaintiff must show:

1. Relevant portions of the work are subject to copyright protection
 2. Plaintiff owns the copyright
 3. Defendant violated Plaintiff's exclusive rights as outlined by 17 U.S.C. §107
- (Fischer, 2017, p. 1).

Plaintiffs sought 1) a statement that “systematic digitization of copyrighted material without authorization was a violation of Sections 106 and 108 of the U.S. Copyright Act; 2) an injunction to prevent the reproduction, distribution, or display of Plaintiffs’ or other copyrighted works except allowed under Section 108; 3) an injunction to prohibit Defendants’ provision of works to Google for digitization without authorization, and 4) the impoundment of all unauthorized digital copies within Defendants’ possession” (*Authors Guild v. HathiTrust*, 2012, p. 2).

In order to maintain clarity, Table 1 outlines the parties involved in each lawsuit. It is particularly important to note that, in order for a judge to hear a case, both parties must have legal standing in the matter at law. During proceedings for *Authors Guild v. HathiTrust* (2012), defendants sought dismissal of the Associational plaintiffs on standing grounds to the extent they assert the rights of their members. Defendants argued the Copyright Act does not include associational standing, thus such parties should be precluded from representation of their members (*Authors Guild v. HathiTrust* (2012), p. 3). Ultimately, Judge Baer held that Associational plaintiffs satisfied the requirements for inclusion and that the issues pertaining to the rights of their members merited review.

Table 1. Plaintiffs and defendants in the four cases.

	Plaintiffs	Defendants	Judges
<i>District Court</i> <i>Authors Guild v. HathiTrust</i> (2012)	Authors Guild, Inc.; The Australian Society of Authors Limited, Authors' Licensing and Collecting Society, Union des Écrivains et des Écrivains Québécois, Sveriges Författarförbund, Norsk Faglitterær Forfatter-og Oversetterforening, Trond Andreassen, Pat Cummings, Erik Grundström, Angelo Loukakis, Helge Ronning, Roxana Robinson, André Roy, Jack R. Salamanca, James Shapiro, Danièle Simpson, T.J. Stiles, and Fay Welson,	HathiTrust; Mary Sue Coleman, President of the University of Michigan; Mark G. Yudof, President of the University of California; Kevin Reilly, President of the University of Wisconsin System; Michael McRobbie, President of Indiana University; and Cornell University	Harold Baer, Jr., District Judge
<i>Authors Guild v. Google</i> (2013)	Authors Guild, Inc., Betty Miles, Joseph Goulden, and Jim Bouton, on behalf of themselves and all others similarly situated	Google, Inc.	Denny Chin, Circuit Judge
<i>Court of Appeals</i> <i>Authors Guild v. HathiTrust</i> (2014)	Authors Guild, Inc., Australian Society of Authors Limited, Union Des Ecrivains Et Des Ecrivains Québécois, Angelo Loukakis, Roxana Robinson, Andre Roy, James Shapiro, Daniele Simpson, T.J. Stiles, Fay Weldon, Authors League Fund, Inc., Authors' Licensing and Collecting Society, Sveriges Författarförbund, Norsk Faglitterær Forfatter-Og Oversetterforening, Writers' Union of Canada, Pat Cummings, Erik Grundstrom, Helge Ronning, Jack R. Salamanca,	HathiTrust; Cornell University; Mary Sue Coleman, President, University of Michigan; Janet Napolitano, President, University of California; Raymond W. Cross, President, University of Wisconsin System; Michael McRobbie, President, Indiana University	Walker, Cabranes, Parker, Circuit Judges
<i>Authors Guild v. Google</i> (2015)	Authors Guild, Inc., Betty Miles, Jim Bouton, Joseph Goulden, individually and on behalf of all others similarly situated	Google, Inc.	Leval, Cabranes, Parker, Circuit Judges

The U.S. Copyright Act grants the owner of a copyrighted work several exclusive rights.

Anyone who violates these rights runs the risk of copyright infringement, and the owners of the copyright is able to bring suit against the infringer. Section 106 of the U.S. Code states:

Subject to sections 107 through 122, the owner of copyright under this title has the exclusive rights to do and to authorize any of the following:

- (1) to reproduce the copyrighted work in copies or phonorecords;
- (2) to prepare derivative works based upon the copyrighted work;
- (3) to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;
- (4) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly;
- (5) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and
- (6) in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission (17 U.S.C. § 106).

In order to fulfill copyright's purpose, "to promote the Progress of Science and useful Arts, by securing limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries" (U.S. Const. art. I, § 8), copyright law has statutory limitations on the exclusive rights of the copyright owner, in particular the fair use exemption. Exemptions allow individuals to use a copyrighted work, within limited circumstances, without knowledge of the copyright holder, without obtaining permission from the copyright holder, and without the payment of a license fee.

A Brief History of Fair Use

To understand fair use, it is important to understand the history of §107. On May 31, 1790, the first copyright law was enacted under the new United States Constitution. Modeled on Britain's Statute of Anne, this first U.S. federal copyright law is relatively limited in scope, protecting books, maps, and charts for only 14 years with a renewal period of another 14 years (U.S. Copyright Office, n.d.). It was not until 1841 that *Folsom v. Marsh* (1841) provided the first copyright case involving fair use. The issue was whether publisher Marsh, Capen, and Lyon's verbatim use of Plaintiffs Sparks and publisher Folsum, Wells, and Thurston's protected letters in *The Writings of George Washington* constituted an "act of piracy" under the 1841 copyright law. The court recognized principles that are the foundation for the modern fair use doctrine. The court states:

In short, we must often, in deciding questions of this sort, look to the nature and objects of the selections made, the quantity and value of the materials used, and the degree in which the use may prejudice the sale, or diminish the profits, or supersede the objects, of the original work (*Folsom v. Marsh* (1841))⁶.

Folsom v. Marsh established the four fair use factors as case precedent, but not until the U.S. Copyright Act of 1976 did Congress codify fair use in 17 U.S.C. § 107. In 1994, during *Campbell v. Acuff-Rose* (1994),⁷ the U.S. Supreme Court described fair use as an affirmative defense, meaning the defendant is required to bear the burden of raising and demonstrating that the use was fair and not infringement. *Campbell v. Acuff-Rose* established that the commercial nature of a work does not automatically render a use unfair. The Supreme Court

⁶ *Folsom v. Marsh*, 9 F. Cas. 342 (C.C.D. Mass. 1841).

⁷ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

applied, for the first time, the transformation analysis for a fair use determination.⁸ Since then, fair use has slowly experienced a shift to placing substantial importance on the transformative nature of the use.

According to Hirtle, Hudson, and Kenton (2009), fair use serves as a primary protection of free speech, provides a balance between the interests of copyright holders and users, and provides protections to libraries, archives, and nonprofit educational institutions (p. 89). Fair use is flexible and open-ended, meaning it can be applied to almost any type of use, responding to changes in technology before the statutes and other policy instruments catch up. Because of the flexibility of fair use, however, no strict guidelines exist for hypothesizing how a court may rule if a use is brought to suit. Each fair use is judged on a case-by-case basis. When determining whether a particular use is fair, the Copyright Act requires consideration of four factors:

1. The purpose and character of the use, such as if the use is commercial or for a nonprofit educational purposes
2. The nature of the copyrighted work, such as if the work is mainly factual or fictional
3. The amount and substantiality of the portion used in relation to the copyrighted work as a whole
4. The effect of the use upon the potential market of the copyrighted work (17 U.S.C. § 107).

These factors are not meant to be individually decisive, but rather viewed as a whole. Prior to 1994's *Campbell v. Acuff-Rose Music, Inc.*, fair use was not systematically evaluated, nor did Congress have a guiding procedure. The *Campbell* Court comprehensively analyzed each of the fair use factors, and future courts used *Campbell* as fair use precedent. It is important to

⁸ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

understand that fair use as a litigation defense is affirmative, i.e., the defendant bears the burden of proof for affirmative defenses. In fair use defenses, the defendant must submit evidence to the court negating civil liability, allowing the court to rule in favor of fair use.

Background to *Authors Guild v. HathiTrust* (2012) and *Authors Guild v. Google* (2013)

In 2012, HathiTrust-partnered libraries entered into an agreement with Google to create digital copies of works held by the libraries. In exchange, Google provided a digital copy to the library that provided the hard copy. The library's digital copy included a scanned image file of all pages and a text file of the entire work. The partnered libraries then contributed their copies to HathiTrust to create a "shared digital repository" which contains almost 10 million digital volumes, with approximately 73% under copyright protection (*Authors Guild v. HathiTrust*, 2012, p. 1). Google used its digital copy for Google Books, an online system that allowed users to search the digital copies held by Google and view "snippets" of the books. Plaintiffs alleged twelve unauthorized copies were made in total by the digitization process. The following section systematically examines the four court opinions on each factor of fair use. Lastly, I will give an overall assessment of fair use for each court case.

The Fair Use Defense in the District Court and Court of Appeals Cases

As discussed above, fair use is determined by the consideration of the four factors: *the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect on the market* (17 U.S.C. §107). Throughout *Authors Guild v. HathiTrust* (2012) and *Authors Guild v. Google* (2013), *Campbell v. Acuff-Rose Music, Inc.* was used as precedent for evaluating the four factors and the overall assessment of the fair use defense. Table 2 below briefly outlines the key points of the fair use factors in the District Court decisions and Table 3 outlines the key points of the fair use factors in the Court of Appeals decisions. The

following section will provide more detail related to the decisions of the courts. Each factor will include the opinion assessment of District Court and the Court of Appeals in *Authors Guild v. HathiTrust* and then the opinion assessment of the District Court and the Court of Appeals in *Authors Guild v. Google*.

Table 2. Key points of the District Court cases.

The Four Fair Use Factors	<i>Authors Guild v. HathiTrust</i> (2012)	Finding	<i>Authors Guild v. Google</i> (2014)	Finding
Factor 1: Purpose and Character of the Use	Creation of a full-text searchable database was highly transformative; access for the print-disabled was a valid purpose and “provides a public good”	Strongly favors fair use	Purpose of enabling a search for identification of books; highly transformative; snippet view adds to the highly transformative purpose; Google’s for-profit, commercial status leans towards unfair use	Transformative nature strongly leans in favor of fair use; commercial status leans against fair use
Factor 2: Nature of the Copyrighted Works	76% of works are fiction	Does not affect fair use	Vast majority of books are non-fiction	Favors fair use
Factor 3: The Amount and Substantiality of the Portion Used	Entire copies are necessary to fulfill the purpose of search capabilities and access for print-disabled individuals	Strongly favors fair use	To enable search function, a full copy is required; snippet view does not reveal enough copyrighted material to infringe author’s rights; revealed matter could not function as a substitute	Favors fair use
Factor 4: The Effect on the Potential Market of the Copyrighted Work	Insignificant present-day market for books accessible to the handicapped; transformative use does not cause copyright holder to suffer market harm due to the loss of license fees	Strongly favors fair use	Type of loss of sale will generally occur in relation to interests not protected by the copyright	Strongly favors fair use
Overall Assessment		Strongly favors fair use		Favors fair use

Table 3. Key points of the Court of Appeals cases.

The Four Fair Use Factors	<i>Authors Guild v. HathiTrust</i> (2014)	Finding	<i>Authors Guild v. Google</i> (2015)	Finding
Factor 1: Purpose and Character of the Use	Full-text searchable database was highly transformative; access for the print-disabled was a valid purpose and “provides a public good”, but not transformative	Favors fair use	Search and snippet views are highly transformative; search function is transformative; Google’s for-profit, commercial motivation in Google’s favor	Transformative nature strongly leans in favor of fair use
Factor 2: Nature of the Copyrighted Works	Full-text search “is not dispositive”; works in HDL are copyrighted	Weights against fair use	Vast majority of books are non-fiction	Does not affect fair use
Factor 3: The Amount and Substantiality of the Portion Used	Entire copies are necessary to fulfill full-text search; multiple copies are required for preservation	Favors fair use	To enable search function, a full copy is required; snippet view displays context of search term, does not favor or disfavor fair use	Favors fair use
Factor 4: The Effect on the Potential Market of the Copyrighted Work	Full-text search does not harm existing or potential traditional markets; Plaintiffs unable to identify any specific instance of harm; present markets for print-disabled are insignificant, HDL provides access to ten million volumes	Strongly favors fair use	Snippet view does not harm market of work; acknowledges <i>some</i> loss of sales, but does not outweigh favorable use	Favors fair use
Overall Assessment		Favors fair use		Favors fair use

The following analysis is structured specifically to analyze how the fair use factors are evaluated in these types of cases. I will start with an introduction of the factor and then

proceed to identify how the factor was evaluated in each of the *HathiTrust* and *Google* cases. The courts emphasize viewing fair use as an overall assessment; however, each factor should be understood as it relates to each case.

THE PURPOSE AND CHARACTER OF THE USE

The purpose and character of the use is the first factor and often viewed as one of the more influential factors. The *Campbell* understanding of the secondary use states,

“Under the first of the four 107 factors...the inquiry focuses on whether the new work supersedes the objects’ of the original creation, or whether and to what extent it is controversially ‘transformative,’ altering the original with new expression, meaning, or message. The more transformative the new work, the less will be the significance of the other factors, like commercialism, that may weigh against a finding of fair use” (*Campbell v. Acuff-Rose Music, Inc.*, 1994 at 591).

When evaluating the purpose of a work, a non-profit use is more likely to garner fair use protection than a commercial use (Hirtle et al., 2009). It is important to note that fair use is a consideration of all four factors, so a commercial purpose does not necessarily negate fair use altogether. When a use is highly transformative, fair use protections are even stronger.

The *Authors Guild v. HathiTrust* Cases

Authors Guild v. HathiTrust, 2012

In the district court, Judge Baer determined that HathiTrust’s Digital Library use of copyrighted material satisfied the first factor. HathiTrust’s Digital Library allows search capabilities for researchers and scholars; preserves libraries’ collections during normal circulation, natural disasters, or theft; and provides print-disabled individuals with access to

the collection. The courts have found that, if the new work provides a transformative use by adding something new or a use that serves an entirely new purpose (*Bill Graham Archives v. Dorling Kindersley Ltd.*, (2006)),⁹ the use is likely to be fair. Judge Baer stated the copies in the digital library served an entirely new purpose: they provided search capabilities, which then provided data mining possibilities for research. The digitization of the libraries' collections transforms printed text into searchable data ripe for data and text mining. Books have evolved from words on the page to facilitating research through the snippets available in the HathiTrust Digital Library.

As part of the agreement between HathiTrust Digital Library and Google, Google provided to each library a digital copy of a hard copy, library-owned work. This digital copy was created by Google, Inc., a for-profit, commercial company. The Plaintiffs argued that HathiTrust's status of "nonprofit, educational institution" should not shield it from copyright infringements. Judge Baer stated, however, the Plaintiffs missed the point of the Defendants' use of the digital copies (*Authors Guild v. HathiTrust*, 2012, p. 8). The Copyright Act entitles some rights to copyrighted works, especially for preservation. Section 108 of the Copyright Act allows libraries, archives, and educational institutions to create digital copies for preservation under specific circumstances. Even if the libraries were to purchase replacement copies, they would still not have access to the data mining and search capabilities that digital copies provide. HathiTrust's use of copyrighted works was deemed fair because of its transformative properties.

⁹ *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F. 3d 605 (2d Cir. 2006).

Authors Guild v. HathiTrust, 2014

The Court of Appeals judges determined the full-text search searchable database is “quintessentially transformative use” (*Authors Guild v. HathiTrust*, 2014, p. 6). The court states the Plaintiffs did not provide evidence they write with the purpose of enabling text searches from their books (p. 6). HathiTrust’s word search function, as detailed earlier in the report, provides only the word and page location, so little to no resemblance remains between the original text and the results of the HDL full-text search. The *HathiTrust* court determined if the full-text search function did not supersede the object’s original creation, the use was transformative and weighed in favor of fair use. HDL is not in the market of republishing original works but adds something new to the original with a different purpose and character.

In contrast with the District Court’s findings of this factor related to print-disabled access, the Court of Appeals did not find access to the print-disabled community a transformative use. The Court of Appeals stated, in order for a use to be considered transformative, the use must contribute something new to the original work. By making copyrighted works available in a print-disabled accessible format, the HDL is expanding their audience, but the underlying purpose of the HDL’s use is the same as the author’s original purpose (p. 9). The plaintiffs argued the converted, print-disabled accessible format constituted a derivative work and the Court of Appeals agreed, citing similarities to translated works. Finding transformative use is not, however, absolutely necessary for a finding of fair use. The judges concluded the Chafee Amendment in the Americans with Disabilities Act would allow appropriate accommodations for print-disabled patrons. The Court of Appeals weighed in favor of fair use for factor one.

The Authors Guild v. Google Cases

Authors Guild v. Google, 2013

In the district court, Judge Chin also determined Google's use of copyrighted material was highly transformative. Chin stated in his opinion, "Google Books has become an important tool for libraries and librarians and cite-checkers as it helps identify and find books" (*Authors Guild v. Google*, 2013, p. 6). As in the HathiTrust case, Google Books does not supersede or replace books because it is not a tool used to read books. Instead, it adds value to the original and allows for "the creation of new information, new aesthetics, new insights, and understandings" (p. 21).

Authors Guild v. Google, 2015

The Court of Appeals judges followed a similar procedure to the *Authors Guild v. HathiTrust* (2014) Circuit judges. Circuit judges cited the transformative uses and examples from *Authors Guild v. HathiTrust* (2014), demonstrating the transformative uses derived from the information in the text itself. The Circuit judges recognized transformative use is not absolutely necessary for a finding of fair use, but a transformative use communicates something new and different from the original or expands its utility (*Authors Guild v. Google*, 2015, pg. 17). The Circuit judges found the transformative use of the search and snippet view functions to support a fair use finding. The search function of Google Books was also in support of fair use since the making of the digital copies and the use of those copies was for searching. In order for a search function to work properly, the entire text was needed. The court cited *Perfect 10, Inc. v. Amazon.com, Inc.*, (2007) and *Kelly v. Arriba Soft Corp.*, (2003) as examples of transformative fair use was found even when the entire work was digitally copied, as long as it served a different function from the original (p. 22).

One of main differences between the *HathiTrust* and *Google* cases is Google's use of a snippet view. HathiTrust does not provide any copyrighted material in the word search, only the word and the page number. Google's search function allows the user to read snippets from the book searched. The snippet view provides the user a context for the searched word, but does not reveal enough to infringe the copyright holder's interests. The court asserted the snippet view provided the highly transformative purpose of identifying potentially interesting works to the user.

The *Campbell* court emphasized, "the effect that the commercial or nonprofit character of a work is 'not conclusive' but merely 'a fact to be 'weighed along with the other[s] in fair use decisions'" (p. 25). The Court of Appeals weighed in favor of fair use because of Google's for-profit, commercial enterprise. While Google, a commercial entity, did copy the books, the Court of Appeals ruled the use fair. Since the institution already owned a physical copy of the title, outsourcing the digitization to Google proved fair because of the transformation of print to digital (*Authors Guild v. Google*, 2015, p.11). They state they do not give much weight to the fact that the defendant's use was commercial.

NATURE OF THE COPYRIGHTED WORKS

The second factor considers the qualities and characteristics of the works being used.

Factual, research-based works tend to weigh in favor of fair use more than creative, fictional works. The more creative the work, the more protection it warrants (Crews, 2012).

***The Authors Guild v. HathiTrust* Cases**

Authors Guild v. HathiTrust, 2012

This factor does not usually play a large role in determining fair use; it can, however, give an idea about the makeup of the collection. The Plaintiffs had "identified 116 works they

alleged were unlawfully digitized by the Defendants, and about 76% of the works were identified as fictional” (*Authors Guild v. HathiTrust*, 2012, p. 9). As mentioned previously, HathiTrust has digitized almost 15 million volumes, which includes text, photos, and other materials, over 7 million book titles, and over 5 billion pages stored as 667 terabytes (HathiTrust, 2017c). But because the transformative use, “intended to facilitate key-word searches or access for print-disabled individuals, the second factor is not dispositive” (p. 9). Where a use is transformative, the nature of the copyrighted works is not likely to “separate the fair use sheep from the infringing goats” (*Campbell v. Acuff-Rose*, 510 U.S. at 586 (1994)).

Authors Guild v. HathiTrust, 2014

The Circuit judges ruled against fair use, explaining print-disabled patrons could obtain access to works that qualify for copyright protection. The Circuit judges explicitly stated print-disabled access does not preclude a finding of fair, given the findings of the other factors.

The *Authors Guild v. Google* Cases

Authors Guild v. Google, 2013

Google set out to digitize the collections of the five participating institutions. All the works were published books of varying genres, in- and out-of-print, and both fiction and non-fiction. While Google digitized both in-and out-of-print books, the vast majorities were out-of-print. According to a study done in 2009, ninety-three percent of books within library collections are non-fiction, while just seven percent are fiction (Lavoie & Dempsey, 2009). Most of the digitized books were non-fiction, weighing in favor of fair use.¹⁰

¹⁰ Judge Chin cited *Blanch*, 467 F. 3d at 253; *Bill Graham Archives*, 448 F. 3d at 612; *Castle Rock Entm’t, Inc. v. Carol Publishing Group, Inc.*, 150 F. 3d 132, 142 (2d Cir. 1998). These cases demonstrated that the Second Circuit Court does not give much weight to the fact that the defendant’s use was commercial.

Authors Guild v. Google, 2015

Agreeing with *HathiTrust* (2014), the Circuit judges felt the second factor was not dispositive, commenting that courts have hardly ever found that the second factor played a large role in determining a fair use decision. The Circuit court further stated, when evaluated in isolation, this factor does not influence the case in one way or another.

THE AMOUNT AND SUBSTANTIALITY OF THE PORTION USED

The third factor examines the amount and substantiality of the portion used compared to the copyrighted work as a whole. The greater the amount taken, the less likely the use is fair (Hirtle et al., 2009). Fair use does not provide a specific measure of what can be used or a formula to determine the ratio of work used compared to the whole. When using copyrighted works, it is understood that using less earns fair use favor, borrowing short quotations and paraphrasing provides stronger fair use protection. Using important quotations and passages, however, can weigh against fair use, such as using the “heart of the work.”¹¹

The *Authors Guild v. HathiTrust* Cases

Authors Guild v. HathiTrust, 2012

Sometimes it is necessary to make a copy of an entire work, especially when search functionality is involved. Without the entire work, search capabilities and data mining do not function properly or accurately. The Defendants’ purpose was to facilitate search capabilities and provide access to the digitized works for print-disabled individuals. Plaintiffs argued the digital copy did not need to be retained for search functionality. Judge Baer stated that the maintenance of an electronic copy was necessary to provide access for print-disabled

¹¹ *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539 (1985)

individuals (*Authors Guild v. HathiTrust*, 2012, p. 9). Judge Baer believed the entire digital copy was necessary, so the copying was not excessive.

The Plaintiffs argued HathiTrust maintained four unauthorized copies of the copyrighted works at different locations. The Digital Library's services are offered on two servers. According to the HathiTrust's executive director, "the existence of an identical mirror site allows for balancing the load of user web traffic to avoid overburdening a single site, and each site acts as a back-up of the digital library collection in the event that one site were to cease operation" (*Authors Guild v. HathiTrust*, 2014, p. 7). The court affirmed that four copies are reasonably necessary to facilitate the services HathiTrust provides to the public and to prevent data loss.

Authors Guild v. HathiTrust, 2014

The Court of Appeals agrees with the District Court's ruling, stating it was reasonably necessary for the HDL to make use of the entire work in order to enable full-text search function. Circuit judges did not find copying to be excessive. As mentioned in the District Court's case, Plaintiffs maintained copies at four locations was excessive, but once again, the Circuit court upheld that these copies were necessary in order to facilitate HDL's legitimate uses. Circuit judges even explain the copies provide data risk prevention and preservation.

The *Authors Guild v. Google* Cases

Authors Guild v. Google, 2013

Google makes two entire scans of the copyrighted work, one for Google's indexing and the other for the lending institution. The entire work needs to be scanned in order to properly index the work and create effective search results. Fair use has been found justified when

complete, unchanged copying is necessary for the transformative properties of the use to be achieved and was done in such a way as to not create a competing substitute for the original. Previous cases such as *Sony Corp. of America v. Universal City Studio, Inc.* provided justification for Google's full-text scan.¹² Whole text digitization is necessary for the transformative properties of search and indexing to function properly. While Google *makes* an entire scan of the work, the public's inability to access the full work plays a role in the fair use evaluation.

Authors Guild v. Google, 2015

Court of Appeals specifically identifies the search function and snippet view when evaluating the third factor. Circuit judges state complete, unchanging copying has repeatedly been found justified, as fair use when copying was reasonably appropriate to achieve the transformative purpose and was done in such a way that did not offer a competing substitute. The Circuit court agreed with the reasoning in *HathiTrust* (2014), "it was reasonably necessary to make an entire copy in order to enable full-text search functions" (*Authors Guild v. Google*, 2015, p. 30). Importantly, when Google makes the digital copy, it does not provide public access. The copy reveals the limited amount allowed by the snippet view.

Google's Snippet View

Upon receiving the search results, the user is not given the entire work, only the snippets related to the search query. The public does not have access to the full digital copy. Plaintiffs felt the user could "cobble" together the snippets to recreate the work, but the court felt the snippet view did not reveal enough information to substitute for purchasing the book. Google constructed the snippet view in a manner that protects against creating a substitute for the Plaintiffs' books. Users were not able to access more than sixteen percent

¹²*Sony Corp. of America v. Universal City Studio, Inc.* (464 US 417, 104 S. Ct. 774, 78 L. Ed. 2d 574-Supreme Court, 1984) agreed that copying the entirety of a work is sometimes necessary for viewing, such as when taping a previously recorded TV program.

of the text through the randomly queried snippets (*Authors Guild v. Google*, 2015). Snippet view is not available for some books, such as dictionaries and cookbooks, where the searched term could reveal the full information needed. Plaintiffs cited *Infinity Broadcast Corporation v. Kirkwood* (1998) and *United States v. American Society of Composers, Authors, and Publishers* (2009) stating, “the defendant’s use that replaces a comparable service license by the copyright holder, even without charge, may cause market harm.” The court found no merit in this argument; the court stated the copyright that protects the Plaintiffs’ works does not include an exclusive derivative right to supply information about the book through query of a digitized copy (*Authors Guild v. Google*, 2015). Judge Chin ruled slightly against fair use. The 2nd Circuit Court of Appeals, however, ruled strongly for fair use, citing the protections of snippet view.

THE EFFECT ON THE POTENTIAL MARKET OR VALUE OF THE COPYRIGHTED WORK

The fourth factor is often viewed as the most influential in determining fair use. When books are copyrighted and in print, they warrant the most protection. The fourth factor focuses on whether the copy creates a complete substitute for the original or its derivative, depriving the rights holder significant revenues because potential purchasers are opting for the substitute. *Campbell* stressed the connection between the first and fourth factors, “in that the more the copying is done to achieve a purpose that differs from the purpose of the original, the less likely it is that the copy will serve as satisfactory substitute for the original” (*Campbell v. Acuff-Rose Music, Inc*, 1994, p. 5). Because copyright was created to protect the creativity of works and allow creators to earn money and create a market for themselves, the fourth factor is heavily scrutinized in fair use cases.

The *Authors Guild v. HathiTrust* Cases

Authors Guild v. HathiTrust, 2012

In this case, the Court stated it considered only those markets the copyright holders were a part of or those they may be willing to develop. Judge Baer cited *Sony Corp. of America v. Universal City Studios, Inc.* (1984), stating that, in the case of a noncommercial, transformative work, the plaintiff must show “by a preponderance of the evidence that some meaningful likelihood of future harm exists” (p. 9). Plaintiffs alleged the digital copies were considered a lost sale; they were created rather than purchased. Judge Baer further explains that a purchased copy would not have allowed either full-text searches or print-disabled access, referring to the first factor that weighed in favor of a finding of fair use based on transformative uses.

Plaintiffs also alleged that HathiTrust, by virtue of having a digital copy of the Plaintiffs’ works, was subject to hacking and theft. Defendants indicated the Center for Research Libraries certified their security as trustworthy and reliable. This argument failed to show infringement, and the “future possibility” of infringement was not enough to constitute further examination.

Because the fourth factor relates to the market for copyrighted works, mainly book sales in the Plaintiffs’ cases, Plaintiffs argued the Defendant’s activities undermine the existing and emerging licensing opportunities, such as a collective management system, which would permit particular activities by the Defendants and provide compensation to the Plaintiffs (*Authors Guild v. HathiTrust*, 2012, p. 10). The Plaintiffs admitted that they could not point to an instance of specific, quantifiable past harm (p. 10). Because of the transformative nature

of HathiTrust's use of the copyrighted works, it's difficult for a copyright owner to anticipate a potential, transformative market. Access for print-disabled patrons does not significantly affect a market, especially since the American Disabilities Act requires providing equitable access as one of the major aspects of this legislation, designed to protect this minority market. Judge Baer states, "a use that falls within a transformative market does not cause the copyright holder to 'suffer market harm due to the loss of license fees'" (p. 10).

Authors Guild v. HathiTrust, 2014

The Court of Appeals agrees with the findings of the District Court. HathiTrust stated the full-text search did not harm any existing or potential markets for the Plaintiffs. The Circuit judges reiterated the lack of evidence on behalf of the Plaintiffs to prove "any specific, quantifiable past harm, or any documents relating to any such past harm" (*Authors Guild v. HathiTrust*, 2014, p. 8). The judges state the "theory" of market harm does support this factor, because the full-text search function does not serve as substitute for the original work. When confronted with the possibility of a data breach, the Circuit judges found no basis in the record to conclude the possibility of a data breach is likely to occur. Plaintiffs' allegations were based on speculation, and the Circuit judges stated a risk of future harm must be "certainly impending," rather than just speculative. The Court of Appeals agreed with the District Court's findings of fair use related to the fourth factor.

The *Authors Guild v. Google* Cases

Authors Guild v. Google, 2013

The Google Books project was not designed to provide substitutes for copyrighted books, only to index them. Indexing increases a work's searchability and discoverability. Google does not sell the scans, and the digitized scans do not replace the books. Essentially, Google is creating an online catalog; it is not in the market of e-book or full textbook online sales

(*Authors Guild v. Google*, 2013). One of the key concerns with this factor was the use of snippet views. Even if the use was highly transformative and was favored by the first factor, did the snippet view nonetheless create a substitute for the original work? The courts found that, in its current construction with access *at maximum* sixteen percent, snippet view could not and did not negatively affect the market. Google Books actually guides users to legitimate purchasing platforms, thus increasing sales of those titles (Brown & Babaei, 2014; Proskine, 2006). “In this day and age of online shopping, there can be no doubt but that Google Books improves books sales” (*Authors Guild v. Google*, 2013, p. 25).

Authors Guild v. Google, 2015

The Circuit court addressed the allegation of derivative rights in the search and snippet view functions; Plaintiffs alleged the snippet views functioned as a derivative work. Plaintiffs cited *Infinity Broadcast Corporation v. Kirkwood* (1998) and *United States v. American Society of Composers, Authors, and Publishers* (2009) stating, “the defendant’s use that replaces a comparable service license by the copyright holder, even without charge, may cause market harm.” The court found no merit in this argument; the court stated the copyright that protects the Plaintiffs’ works does not include an exclusive derivative right to supply information about the book through query of a digitized copy (*Authors Guild v. Google*, 2015). Also, since the institution already owned a physical copy of the title, the transformation of it into the digital realm for the purpose of search ability does constitute fair use, even if they “outsourced” the digitization to Google (p. 11).

THE OVERALL ASSESSMENT

When fair use evaluations are conducted, the overall assessment decides the fairness. While each factor is evaluated individually, a slightly less fair factor can be outweighed by a

stronger fair use factor. All factors are weighed together, along with any other relevant considerations pertaining to copyright law.

The Authors Guild v. HathiTrust Cases

After weighing each of the four factors and reflecting upon the main purpose of copyright, “to promote Science and the useful Arts,” Judge Baer ruled in the favor of HathiTrust and found their use fair. The transformative nature that allowed print-disabled patrons access to materials, the preservation capabilities, and the fact that no in-copyright material was made fully available heavily weighed in the favor of fair use. HathiTrust supported scholarship beyond just a digital repository, and the access the repository provides to print-disabled students was previously impossible, because of financial and staffing constraints.

“Defendants’ mass digitization project would require that I terminate this invaluable contribution to the progress of science and cultivation of the arts that at the same time effectuates the ideals espoused by the ADA” (Judge Baer, *Authors Guild v. HathiTrust*, 2012, p. 11).

The Circuit Court of Appeals generally agreed with the rulings delivered by Judge Baer. While the nature of the copyrighted works is not the most important factor in the assessment, the two courts disagreed on this factor. The second factor did not have any weight in the findings of fair use for the District Court. The Circuit court, however, ruled slightly against a fair use finding. Circuit judges stated the works digitized, by merit, these works at issue are the type that the law values and seeks to protect. But the second factor is not enough to sway the finding in a particular direction. The first factor is the main point of disagreement, but not enough to alter the overall finding of fair use. In the first factor, the Circuit judges did not agree with Baer’s transformative assessment of the availability of the

works for the print-disabled. Their assessment determined that these works were essentially derivative, such as a translation of the work to a different language, only in this case, a different file format. Overall, the Circuit Court upheld the District Court's findings of fair use.

The *Authors Guild v. Google* Cases

After weighing each fair use factor separately, Google's fair use defense was upheld by Judge Chin. The Google Books project provides significant public and research benefits. The project expands discoverability and access to information for many users, previously not possible. It is free from infringement and will drive users to legitimate book sales, increasing authors' and publishers' revenue and continually expanding their audiences. Additionally, Google Book Search serves as a preservation tool, particularly for out-of-print and older books. It facilitates access to books for print-disabled and remote or underserved populations. Judge Chin fully supported Google's defense of fair use in its massive Google Books Search and Google Library Project.

The Circuit Court of Appeals generally agreed with the overall findings of fair use. Both courts found the highly transformative nature of the use to be the strongest force behind the decision. Only minor differences characterized the opinions. Judge Chin of the District Court did find against Google's for-profit corporation status. The Circuit judges held this aspect as a positive factor for fair use, where the *Authors Guild v. HathiTrust* cases, found the commercial status unfair. Both courts agreed the transformative nature of the snippet view supported a fair use finding concerning market value. The Circuit Court admits the 16% availability through the snippet view *could* lead to potential loss of revenue, but the court

finds this loss does not suffice to make the copy an effective competing substitute that would sway the fair use findings. Overall, the Court of Appeals Google's Book Library's use fair.

Conclusion

In this section, I provided the foundational framework to understand and evaluate a copyright case. As stated earlier, this section aimed to explain the various elements of copyright law implicated by MDPs of copyrighted and public domain works, focusing heavily on the statutory exemption of fair use. By examining the *Authors Guild v. HathiTrust* and *Authors Guild v. Google* cases, the decisions demonstrate how fair use applies in the context of mass digitization projects within libraries. By analyzing each of the factors of fair use, the main similarities of the two cases begin to emerge. In the final section of this report, I would like to provide suggestions for libraries in the planning and development stages of their digitization projects. I have outlined how fair use is determined in the judicial system and will now turn to the structure of HathiTrust in relation to the court decisions and how HathiTrust can provide a model for other libraries' digitization projects.

CONSIDERATIONS FOR MASS DIGITIZATION PROJECTS WITHIN LIBRARIES

With the HathiTrust Digital Library upheld in both the District Court and the Court of Appeals, what do the decisions really mean for libraries? Judge Chin stated, “if there’s no liability for copyright infringement on the part of the library, then there cannot be a liability on the part of Google” (*Authors Guild v. Google*, 2013, p. 27). The HathiTrust and Google decisions demonstrate how fair use applies in the context of digitization projects within libraries and provides a guide for a library looking to digitize their analog works collections. “The underlying logic for university libraries to participate with Google was the same reason universities prevailed in the litigation,” said a HathiTrust staff member. This section aims to provide guidelines highlighted from the court cases that played a direct role in the affirmation of fair use. As libraries continue to navigate the procedures of establishing digital libraries, they may wish to incorporate some of the following suggestions into their digital library development. Additionally, it is essential to follow the advice of legal counsel and of professional associations.

Essential Digital Library Characteristics for a Fair Use Defense

As addressed in the previous section, in order to earn an overall finding of fair use, a digital library project must maintain certain characteristics, such as search ability and limited impact on the market. Judges Baer and Chin of the District Court and the Circuit Judges stated in their opinions that the advanced search capabilities did not reveal enough in-copyright material to damage rightsholders’ interests. HathiTrust’s search function did not reveal any copyrighted material. Google Books, however, provided a snippet view, determined transformative. The characteristics of a transformative use within the context of a digitization project will be discussed later in this section.

Providing a level of public good may not make a use transformative, but the Court of Appeals and Judge Baer explicitly state a use that provides a public good is protected under fair use. The Court of Appeals in *HathiTrust* stated that access to copyright materials by print-disabled patrons does not qualify as a transformative, but it functions as a derivative work, such as a language translation. It is worth mentioning that libraries and archives considering digitization projects automatically earn a non-profit, educational status. As illustrated in the previous section, since *Campbell*, evaluation of the commercial aspects of a use are not considered conclusive, but weighed along with the others in a fair use decision.

The Transformative Nature of the Digital Library

Since *Campbell v. Acuff-Rose* (1994), the judicial interpretations of the first factor of fair use has evolved from an assessment of for-profit/not-for-profit uses into an understanding that the defendant's use focuses on whether the new work "supersedes the objects' of the original creation ...or instead adds something new, with a further purpose." Transformative uses provided the strongest basis for a finding of fair use in the four cases discussed above. The transformative property HathiTrust is promoting moves beyond the traditional, creative approaches to a transformative use, just as 2 Live Crew's "Pretty Woman" parody or Richard Prince's work transformed Patrick Cariou's photograph. HathiTrust and Google's Book Library are non-expressive, technology-oriented transformations of creative works. As with any fair use analysis, inherent uncertainty can leave libraries and archives feeling vulnerable about their digitization efforts. Digitization efforts mean new possibilities of scholarship by students and faculty using these transformed tools. Naturally, it is imperative to consult legal counsel before starting any digitization. Through the Circuit courts' decisions, however, it is affirmed and evident that a searchable, digital database is a strong

transformative use. The Circuit Court's remand of Judge Baer's identification of HathiTrust as a public good may have missed the point he was trying to make, that HathiTrust is creating contributory transformative use.

Changes to HathiTrust after the Litigation

As part of the opinion and ruling, a judge may order an injunction to either stop or change aspects of the infringing use in an attempt to stop the infringement. During an interview with a high level member of the HathiTrust staff, I asked if any changes had been made to any policies or procedures of HathiTrust. HathiTrust did not need to make any changes, as the court did not order an injunction. HathiTrust took it upon them to review their own policies, "we were doing what was agreed, our activity has not changed and we have more comfort and confidence," said the HathiTrust staff member. Later, it was shared that some university partners only provide public domain items. As a way of complying with copyright when the copyright status of a digital object is not entirely clear, HathiTrust restricts access to any material not in the public domain.

Joining HathiTrust and the Copyright Review Management System Toolkit

The University of Michigan Library received a grant to develop a system toolkit for other libraries looking to digitize their materials. The toolkit was designed to help libraries responsibly identify and provide access to their own public domain materials. The toolkit provides planning documents used for the early stages of development, practical considerations of copyright for the project, and reports on pilot projects to provide a sense of opportunity and limitation to the projects. During the interview, it was stressed that while the justice system has upheld HathiTrust and Google, it does not mean libraries can be reckless with their digitization and access. "We continue to be stewards of the files, but we

are also stewards of copyright and privacy,” stressed the HathiTrust member, “it is imperative we provide sophistication and effort in doing this well, by being part of a consortium, we are doing this in a good way.”

One piece of advice provided from the interview suggested libraries consider joining HathiTrust instead of building their own digitization project. “If a library does not have a strong handle, plenty of training for staff, and access to legal council with specific knowledge of advocating for rightsholders, it can hurt the trust we have gained with these projects,” said the staff member. For a library looking to start a project, they may consider joining HathiTrust and sharing the expertise and infrastructure that has already been developed. As part of the litigation, the Plaintiffs stressed concern for the security of the files. A smaller institution needs departmental resources to keep everything secure or consider working with an institution with those resources, such as HathiTrust. By following the CRMS toolkit, or joining HathiTrust, or a combination of both, libraries can discuss the issues of their collection, make informed decisions, and take pieces of each option that works best for their institution.

Reproductions for Use and Preservation

As stated earlier, HathiTrust was developed to address the brittle books issues libraries across the world are facing. The need for preservation allowed HathiTrust to develop a strategy to make their collections accessible. During the litigation, the Plaintiffs in *Authors Guild v. HathiTrust* (2012) and *Authors Guild v. Google* (2014) established a *prima facie* case of infringement. Based on first impression, both digital collections appeared to be blatant reproductions and distributions of in-copyright material. As seen through the examination of the four factors of fair use, however, the search functions of HathiTrust do not reveal any

in-copyright material, and Google allows view of only about 16% of copyrighted material through the snippet view when the user searches the database.

The multiple copies created by Google and HathiTrust were a major concern in the Authors Guilds' argument. Plaintiffs alleged that the multiple copies created by Google for the libraries were unnecessary and unauthorized. The court, however, considered them reasonably necessary for the HathiTrust Digital Library in order to enable full-text search functions (*Authors Guild v. HathiTrust*, 2012). The court further said that the four copies made for the library were justified for the purpose of balancing the load of web traffic and as a back-up in the case of a disaster (Band, 2015). The preservation copies were not available as versions made available. But, by storing the digital copies of the books, HathiTrust preserves them for generations to come, long after the copyright term expires. The HathiTrust decision indicates that a library looking to start its own digitization project could reasonably make digital copies of the print materials in its collection and store those copies, if the library provided full-text search capabilities and full-text access for print-disabled patrons. In addition to following a fair use analysis, libraries and archives can augment their fair use defense with the protections awarded under 17 U.S.C. §108.

There is no general provision in copyright law that permits libraries and archives to reproduce and preserve published material (Hirtle, Hudson, and Kenyon, 2009). Section 108 of the U.S. Copyright Act, however, does allow libraries and archives to reproduce collection items in certain specific instances without payment or permission from the copyright holder. As Hirtle, Hudson, and Kenyon (2009) state, §108 provides some "bright line" rules that exempt certain types of reproduction from infringing use, as opposed to the uncertainty

inherent in fair-use analyses. Section 108 even authorized some reproductions that would be found infringing under fair use analysis, such as whole copying for preservation purposes. Section 108 can be useful for libraries looking to employ multiple protections for their digitization project.

Access for Print-Disabled Patrons

The unique characteristic of HathiTrust as opposed to Google's Books Library is HathiTrust's dissemination of full-text works to print-disabled patrons. The copies created by Google and made available in HathiTrust and the facilitated access for print-disabled patrons has been established as transformative. Judge Baer's analysis of the first factor of fair use indicated print-disabled individuals are not considered a significant market to the Plaintiffs. Regardless of its transformative use, an access copy made for a disabled patron falls within the guidelines for fair use (*Authors Guild v. HathiTrust*, 2012). While the court was not specific about which disabled persons are entitled to full-text access, discussion of the various digital files indicates that both blind and print-disabled patrons, i.e., those persons whose physical impairments prevent them from turning pages or holding books, are entitled to use the full-text and image files of a copy. The transformative nature of HathiTrust provides more equitable access to materials for disabled patrons, further advancing the Americans with Disabilities Act.

The 1990 Americans with Disabilities Act includes a provision for access to copyrighted materials for disabled users. Since print-disabled versions are not considered a profitable market, however, many copyrighted materials are not made available to print-disabled patrons. The provision of equal access to copyrighted material of the ADA was included based on the historical disadvantage disabled patrons have incurred. The *U.S. Code* states that

the ADA was designed to provide the “opportunity to compete on an equal basis and to pursue those opportunities for which our free society is justifiably famous” (42 U.S.C. § 12101(a)(8)). Under the Chafee Amendment to the Copyright Act, authorized entities are permitted to make reproductions of copyrighted material in specialized formats exclusively for use by print-disabled patrons (17 U.S.C. § 121). Currently, the University of Michigan is the only entity that has made its works available to print-disabled students. Judge Baer indicated those HathiTrust Digital Library would fall within the authorizations allotted by the Chafee Amendment, not needing to rely on only a fair use defense. As highlighted from Judge Baer’s District Court opinion:

I cannot imagine a definition of fair use that would not encompass the transformative uses made by the Defendants’ MDP [mass digitization project] and would require that I terminate this invaluable contribution to the progress of science and cultivation of the arts that at the same time effectuates the ideals espoused by the ADA (*Authors Guild v. HathiTrust*, 2012).

Even though the Circuit Court found that HathiTrust’s access to materials for print-disabled persons is a misapprehension of transformative use, the judges concluded providing access to print-disabled patrons fulfills a valid purpose under the first factor, even if such use is not transformative. During the *Sony Corp. of Am.*, Judge Stevens stated, the House Committee Report, a document accompanying a measure reported from a committee, expressly identified making a copy of a copyrighted for the convenience of a disabled patron as an example of fair use (*Authors Guild v. HathiTrust*, 2014, p. 10). For libraries, even if a copy does not immediately present transformative properties, other instances of fair use may be available.

Conclusion

In this section of the report, we have addressed and discussed the HathiTrust and Google decisions and how fair use applies in the context of digitization projects within libraries, beyond transformative use. If libraries are making preservation copies or digitizing works for disabled community members, such actions may not be considered transformative. By applying fair use differently or different sections of copyright law, however, libraries may legally be able to digitize their analog collections. If a library is undertaking a digitization project, it is important to evaluate the risks versus the rewards. By following the Copyright Review Management System toolkit, libraries can make informed decisions and ask the necessary questions early on in project development. Digitizing a collection that may contain public domain and copyrighted works may run the risk of receiving take down notices from copyright holders. Some libraries, however, may view the benefits of access and preservation outweighs the risks.

CONCLUSION

The U.S. Copyright Act is designed with the “goal of promoting the Progress of Science and the useful Arts” by allowing creators and other rightsholders to financially benefit from their works for a limited amount of time. This report aims to understand how a mass digitization project maintains compliance with current copyright law, particularly the fair use clause in copyright law (17 U.S.C. § 107).

This research attempted to ask: How has consideration of fair use played a role in the structure of HathiTrust? How can libraries learn from the four cases discussed and develop a digitization framework that allows for a strong fair use defense?

After examining the infrastructure of HathiTrust and the District Court and Court of Appeals rulings for both HathiTrust and Google cases, researchers and library professionals have a better understanding of how to develop MDPs. Fair use provides an exception to rightsholders’ exclusive rights under copyright, promoting research and scholarship, when such use would be otherwise considered infringement. The HathiTrust and Google decisions have established the courts’ understanding of the implications of mass digitization projects done by libraries and their effect on the public good. The structure and policies employed by HathiTrust have allowed wide information dissemination while maintaining copyright compliance for the objects included within their library. Future libraries can see the decisions made by HathiTrust and the courts concerning fair use and determine the range of access they may wish to allow to their own collections.

The court decisions have also provided some insight into a changing legal landscape. The commercial or for-profit use of a copyrighted work is becoming a secondary consideration for fair use. The merits of these cases were decided within the Southern District Court of New York and the 2nd Circuit Court of Appeals. If copyright owners took action against a library's digitization project, a different judge in a different circuit may interpret the facts of such new cases differently. The court system does not provide certainty.

In addition to the uncertainty of the court system, many copyright scholars are critical of Google's involvement with over 150 libraries' support for scholarship and custody of the cultural record. Should a single, for-profit corporation have this much control over cultural materials? Right now, libraries are embracing the manpower and funding provided by Google, but does Google have too much control of the world's knowledge? What about the information not part of the HathiTrust, Google Books Library, California Digital Library, and others? How can libraries ensure all types of historical and cultural documents are preserved? Some of the challenges HathiTrust faces is its international partners and the international materials these partners provide. A HathiTrust member explained, "we are no longer dealing with U.S. only materials, should we do things for the partners in other countries?" These are just some questions to ask while HathiTrust and Google continue to expand one of the world's largest digital libraries.

Ultimately, this report attempted to provide a legal context for libraries considering their own digitization projects. The HathiTrust infrastructure, particularly the search functionality, accessibility for disabled patrons, rights access determination protocols, the Rights Database, and the metadata for the library's holdings provide a foundation for other libraries to

consider in their own project development. By implementing some of these practices, the digital library's infrastructure contributes to a potential fair use defense before any object is digitized. Further, the four opinions discussed here provide a legal framework for a library's possible fair use defense. Both the courts for each case reached essentially the same decision, HathiTrust's uses were non-infringing; signifying that mass digitization for full-text search and the provision of access to print-disabled patrons are non-infringing and sufficiently transformational uses. These examples also offer the library community justifications for a digital library beyond just transformative, fair use protections, such as reproduction of copies for preservation and providing access to print-disabled users. As with any digitization project involving both public domain and in-copyright materials, some risk is involved. Is the reward is worth the risk?

REFERENCES

29 U.S.C. § 794 (d)

17 U.S.C. §108

17 U.S.C. §107

17 U.S.C. § 106

17 U.S.C. § 121

U.S. Const. art. I, § 8

42 U.S.C. § 12101(a)(8)

A&M Records, Inc. v. Napster, Inc., 239 F. 3d 1004 (9th Cir. 2001)

Authors Guild v. Google, 804 F. 3d 202 (2d Cir. 2015)

Authors Guild v. Google, 954 F. Supp. 2d 282 (S.D.N.Y. 2013)

Authors Guild v. HathiTrust, 755 F. 3d 87 (2d Cir. 2014)

Authors Guild v. HathiTrust, 902 F. Supp. 2d 445 (S.D.N.Y. 2012)

Bill Graham Archives v. Dorling Kindersley Ltd., 448 F. 3d 605 (2d Cir. 2006)

Blanch v. Koons, 467 F. 3d 224 (2d Cir. 2006)

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 114, 114 S. Ct. 1164 (1994)

Castle Rock Entm't, Inc. v. Carol Publishing Group, Inc., 150 F. 3d 132, 142 (2d Cir. 1998)

Folsom v. Marsh, 9 F. Cas. 342 (C.C.D. Mass. 1841)

Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539 (1985)

Infinity Broadcast Corporation v. Kirkwood, 150 F. 3d 104 (2d Cir. 1998)

Kelly v. Arriba Soft Corp., 336 F. 3d 811 (9th Cir. 2003)

Kelly v. Arriba Soft Corp., 336 F. 3d 811, 819 (9th Cir. 2003)

Perfect 10, Inc. v. Amazon.com, Inc., 508 F. 3d 1146, 1165 (9th Cir. 2007)

Sega Enters. Ltd. v. MAPHLA, 948 F. Supp. 923 (N.D. Cal. 1996)

- Sony Corp. of America v. Universal City Studio, Inc.*, 464 US 417, 104 S. Ct. 774, 78 L. Ed. 2d 574- Supreme Court, 1984)
- United States v. American Society of Composers, Authors, and Publishers*, 331 F. 2d 117 (2d Cir. 1964)
- Band, J. (2015). What does the HathiTrust decision mean for libraries? *Research Library Issues: A Report from ARL, CNI, and SPARC*, 285, 7–13.
- Bellardo, T. (2006). Impacts of Mass Digitization Projects on Libraries and Information Policy. *Bulletin of the American Society for Information Science and Technology*, 33(1), 20–24.
- Brown, E., & Babaei, R. (2014). The Authors Guild, Inc. v. Google, Inc.: The Google Books case comes to a close (for now). *Copyright & New Media Law Newsletter*, 18(1), 3–10.
- Crews, K. D. (2012). *Copyright Law for Librarians and Educators* (3rd ed.). Chicago: American Library Association.
- Duffy, E. P. (2013). Searching HathiTrust: Old concepts in a new context. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 8(1).
- Fischer, W. (2017). Steps of a copyright infringement lawsuit. Cambridge, Mass. Retrieved from <https://www.youtube.com/watch?v=kYUloapzo5U&feature=youtu.be>
- HathiTrust. (2016). Google hacks: 100 industrial strength tips and tricks; About this book. Retrieved April 18, 2017, from <https://babel.hathitrust.org/cgi/pt?id=mdp.39015055813896;q1=Google>
- HathiTrust. (2017a). Access Determination for HathiTrust Objects. Retrieved February 3, 2017, from https://www.hathitrust.org/access_determination
- HathiTrust. (2017b). Mission and Goals. Retrieved February 3, 2017, from https://www.hathitrust.org/mission_goals
- HathiTrust. (2017c). Statistics and Visualizations. Retrieved February 3, 2017, from https://www.hathitrust.org/statistics_visualizations
- Hirtle, P., Hudson, E., & Kenton, A. (2009). *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums*. Ithaca, New York: Cornell University Press.
- Lavoie, B., & Dempsey, L. (2009). Beyond 1923: Characteristics of potentially in-copyright print books in library collections. *D-Lib Magazine*, 15(11–12). <http://doi.org/10.1045/november2009-lavoie>
- Levine, M., Adler, R., Bonfiglio, J., Eden, K., & Hall, B. (2016). *Finding the Public Domain: Copyright Review Management System Toolkit*. Ann Arbor: Michigan Publishing.

<http://doi.org/http://dx.doi.org/10.3998/crmstoolkit.14616082.0001.001>

Proskine, E. A. (2006). Google's technicolor dreamcoat: A copyright analysis of the Google Book Search Library Project. *Berkeley Technology Law Journal*, 213–239.
<http://doi.org/10.15779/Z38V69N>

Stanford University Libraries. (n.d.). Welcome to the public domain. Retrieved April 18, 2017, from <http://fairuse.stanford.edu/overview/public-domain/>

Supreme Court of the United States. (n.d.). Frequently asked questions. Retrieved April 18, 2017, from <https://www.supremecourt.gov/faq.aspx>

U.S. Copyright Office. (n.d.). History and education. Retrieved April 18, 2017, from <https://www.copyright.gov/history/>

United States Courts. (2015). *U.S. District Courts-Copyright, Patent, and Trademark Cases Filed, by Jurisdiction*. Retrieved from <http://www.uscourts.gov/statistics/table/47/judicial-facts-and-figures/2015/09/30>